Big Data & Cyber Security

Technica®

What are we trying to protect?

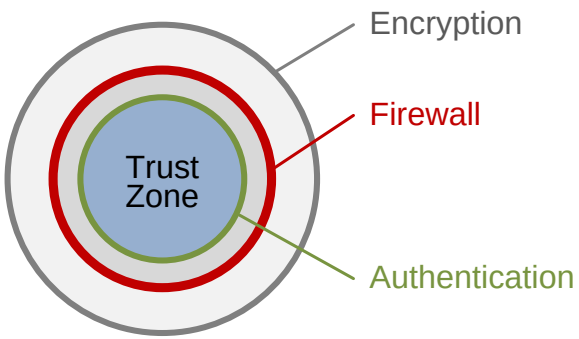
People

Systems

Data

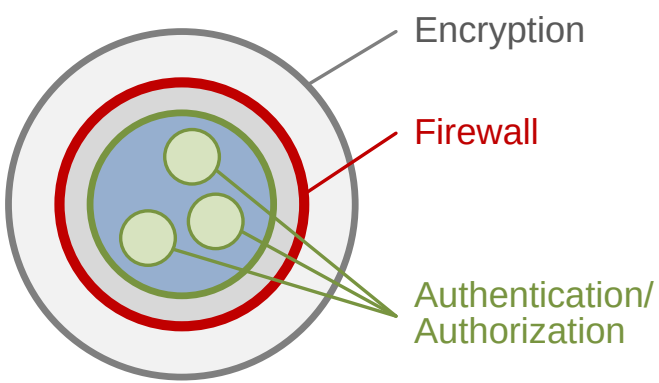
How do we protect?

Traditional



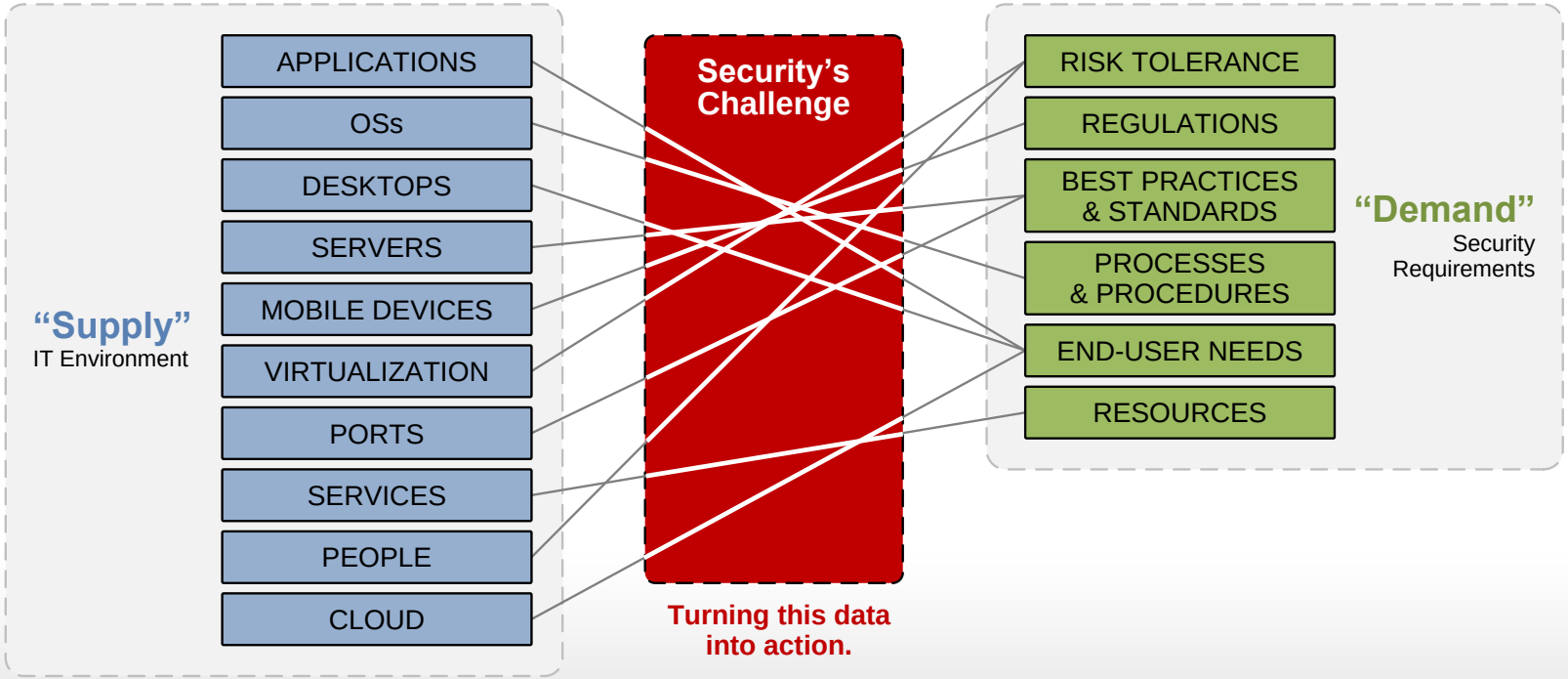
- Keep data secret between sites.
- Control access.
- Trust on the inside.

Zero Trust Model (Forrester, 2010)



- All resources accessed in a secure manner.
- Access control on a need to know basis, strictly enforced.
- Systems must verify, never trust.
- Must inspect and log all traffic so someone can review it.

There is no shortage of security data…



Quote from Cloud Security Alliance

"Enterprises routinely collect terabytes of security relevant data (e.g., network events, software application events, and people action events) for several reasons, including the need for regulatory compliance and post-hoc forensic analysis. Unfortunately, this volume of data quickly becomes overwhelming. Enterprises can barely store the data, much less do anything useful with it."

- Cloud Security Alliance, Big Data for Security Analytics, 2013

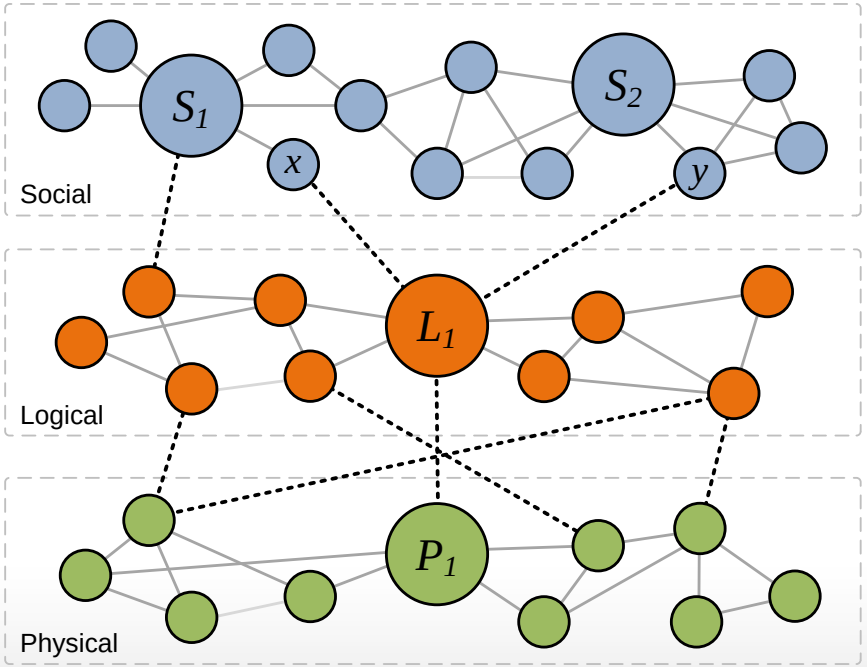
Technica®

How can we approach Cyber Security with Big Data Analytics?

- Real Time Protection:
  - Intrusion Detection Systems
  - Intrusion Protection Systems
  - Complex Event Processing
- Strategic Protection (Big Data Analytics):
  - Graph Analytics
  - Classification Algorithms
  - Abnormal State Detection

Technica®

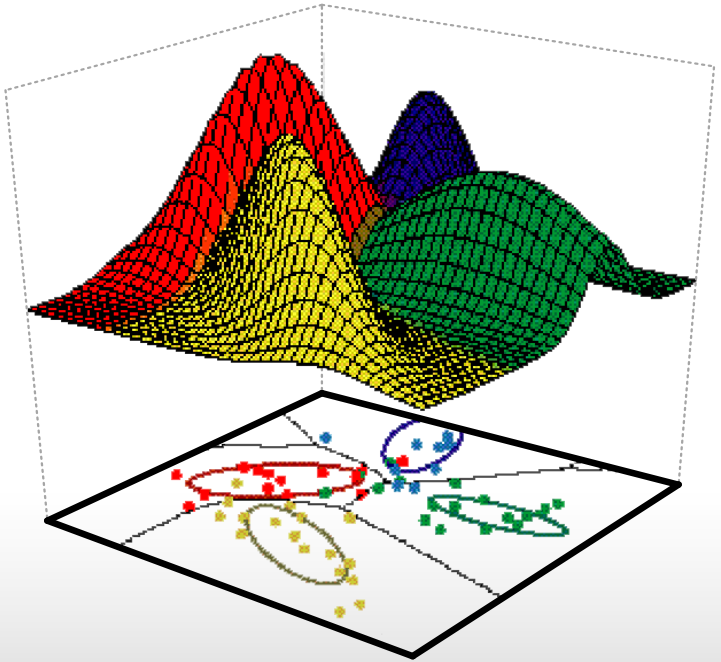
Graph Analytics



- PageRank:
  - Which nodes have the most connections?
- Belief Propagation:
  - Where might the next incident occur?
- Community Detection
  - What is the behavior of the connections?
  - Are they growing?
  - Decreasing?

Classification Algorithms

- Rules can be created from the raw data gathered using different Classification Algorithms:
  - Artificial Neural Networks (ANN)
  - Decision Trees (DT)
  - Naive Bayes (NB)
  - Support Vector Machines (SVM)
  - Boosted Decision Trees (BDT)
  - Boosted Naive Bayes (BNB)

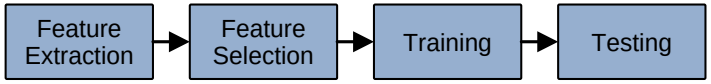


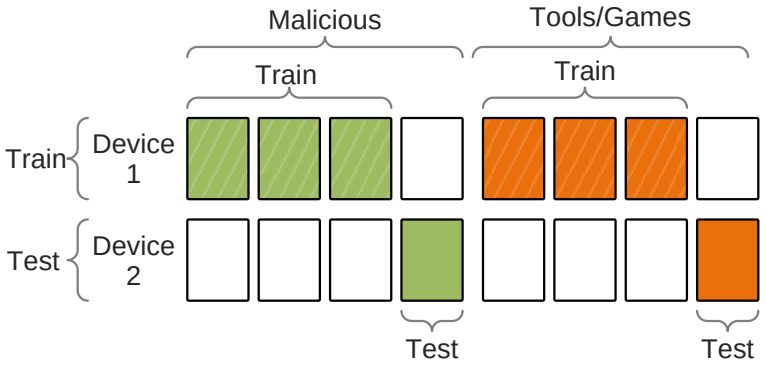
CYBERVIRGINIA

Technica®

Abnormal State Detection

- Identify the most informative features to monitor.
- Evaluating various detection methods and algorithms.
- Understanding the feasibility of running these methods for detection.





- Detection Algorithms: K-Means, Histograms, Logistic Regression, Decision Tree, Bayesian Net, Naive Bayes
- Feature Selection: InfoGain, Chi Square, Fisher Score
- Differentiate between applications which are not included in the training set when training and testing are performed on different devices.

What type of question do you want to answer?

Predict Discrete Attributes

Algorithms	Missions
▪ Collaborative Filtering ▪ K-Means ▪ Principal Component Analysis ▪ Belief Propagation	▪ Determine which entry ports are of most interest to a given threat ▪ Determine type of threat based on specific activity ▪ Infer an individual's tendencies based on those of his friends and family

Predict Continuous Attributes

Algorithms	Missions
▪ Collaborative Filtering	▪ Predict site visitors given historical trends ▪ Predict how an insider threat might value certain risk factors ▪ Predict likelihood that a packet might contain malware items

Determine Groups

Algorithms	Missions
▪ Community Detection ▪ K-Means ▪ Belief Propagation	▪ Analyze individuals by patterns ▪ Identify servers with similar usage characteristics ▪ Determine groups persuaded by similar interests

Predicting Influencers

Algorithms	Missions
▪ Page Rank ▪ Community Detection	▪ Determine group dynamics based on link analysis ▪ Determine the most efficient message dissemination

Technica®

There are challenges

Storage	Cost vs. Efficiency	Expertise
▪ Where and what format? ▪ Where you store your data effects how your solution accesses it. ▪ The format of your data effects the efficiency of preparing the data for analysis.	▪ Performing computations on terabytes and petabytes of data takes hours/day/weeks on traditional systems. ▪ Clusters and high end servers can overcome a budget.	▪ We need more Data Scientists and Algorithm Developers just as we need more Cyber Security Experts.

Technica®

Don’t forget: Secure your Big Data

Big Data Security & Privacy Challenge	
▪ Secure Computations in Distributed Programming Frameworks ▪ Security Best Practices for Non-Relational Data Stores ▪ Secure Data Storage and Transactions Logs ▪ End-Point Input Validation/Filtering	▪ Real-Time Security Monitoring ▪ Scalable and Composable Privacy-Preserving Data Mining and Analytics ▪ Cryptographically Enforced Data-Centric Security ▪ Granular Access Control & Audits ▪ Data Provenance

- Cloud Security Alliance, 2013

Technica®

Conclusion

- We have tools to protect our resources.
- We have ‘Big Data’ that might help us protect our resources more reliably.
- Choosing a Tool to fit your budget means knowing the questions you want to ask.
- Don’t forget to protect your Big Data resources too!

Technica®

Point of Contact

Rhonda Eldridge

Director of Engineering

e: reldridge@technicacorp.com

p: 703.662.2124

Technica Corporation
22970 Indian Creek Dr., Suite 500
Dulles, VA 20166
703.662.2000

technicacorp.com

Rhonda Eldridge has over 24 years of professional experience working in Virginia pioneering Research & Development. As Technica Corporation's Director of Engineering, Ms. Eldridge has responsibility for internal research and development, visioning, and business development -- focusing on cutting-edge cyber security and IT projects for Federal customers, including DoD and IC. She is a communications engineer with strong educational credentials and substantial experience managing projects that utilize the latest technologies. Her work has spanned the technology lifecycle including research and development, laboratory testing, implementation, and ongoing systems support. Ms. Eldridge has supervised and implemented numerous IT projects, including Cyber Security projects, cloud services projects, Service Oriented Architecture (SOA)-based software implementation projects, optical equipment installations, and large-scale Internet Protocol (IP) network implementations.

CYBERVIRGINIA

Technica®